


| | |
|---------------------------------------------------------------------------------------------------------------|------------------------------|
|  (Enriching the Research) | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

Predicting the Costumer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques

¹B Mayuka Reddy, ²Pranavi Nyalapatlawar, ³M Pravalika, ⁴K Joseph Thanusha

^{1,2,3}B.Tech Students, ⁴Assistant Professor

Department of Electronics and Communication Engineering

Bhoj Reddy Engineering College for Women, Saidabad, Hyderabad

Abstract


Customer churn analysis and prediction in the telecom sector is an issue nowadays because it's very important for telecommunication industries to analyse the behaviours of the various customers to predict which customers are about to leave the subscription from telecom companies. So, data mining techniques and algorithms play an important role for companies in today's commercial conditions because gaining a new customer's cost is more than retaining the existing ones. In this paper, we can focus on various machine learning techniques for predicting customer churn through which we can build the classification models such as Logistic Regression, SVM, Random Forest, and Gradient boosted tree and Light GBM and also compare the performance of these models.

I.INTRODUCTION

In today's technological conditions, new data are being produced by different sources in many sectors. However, it is not possible to extract the useful information hidden in these data sets, unless they are processed properly. To find out this hidden information, various analyses should be performed using data mining, which consists of numerous methods. The Churn Analysis aims to predict customers who are going to stop using a product or service among the customers. And, the customer churn analysis is a data mining-based work that will extract these possibilities. Today's competitive conditions led to numerous companies selling the same product at quite a similar service and product quality. With the Churn Analysis, it is possible to precisely predict the customers who are going to stop using services or products by assigning a probability to each customer. This analysis can be performed according to customer segments and the amount of loss (monetary equivalent). Following these analyses, communication with the customers can be improved to persuade the customers and increase customer loyalty. Effective marketing campaigns for target customers can be created by calculating the churn rate or customer attrition. In this way, profitability can be increased significantly or the possible damage due to customer loss can be reduced at the same rate. For example, if a service provider which has a total of 2 million subscribers, gains 750.000 new subscribers and lost 275.000 customers; the churn rate is calculated as 10%. The customer churn rate has a significant effect on the financial market value of the company. So, most of the companies keep an eye on the value of the customer during monthly or quarterly periods.

1.1 Motivation

From the problems obligatory through market saturation and value implications, there has been associate identification of a desire for a laptop based mostly churn prediction methodology that's capable of accurately distinctive a loss of client ahead, so proactive retention ways is deployed during a bid to retain the client. The churn prediction should be correct as a result of retention ways is pricey. A limitation of current analysis is that alternative studies have focused virtually solely on churn capture, neglecting the problem of misclassification of non-churn as churn. Retention campaigns usually

| | |
|-----------------------------------------------------------------------------------|------------------------------|
|  | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

embrace creating service based mostly offers to customers during a bid to retain them. These offers are pricey, thus providing them to customers World Health Organization don't shall churn will have a substantial impact on the whole price of a retention strategy. an extra limitation of current analysis is that it's typically supported one output within the kind of zero for non-churn and one for churn. This has been recognized as a limitation as a result of it restricts analysis prospects. So as to handle the issues mentioned higher than, a profile based mostly analysis methodology is known as a doable answer.

1.2 Overview

In order to retain existing customers, Telecom providers need to know the reasons of churn, which can be realized through the knowledge extracted from Telecom data. In this project, we train machine learning models which is Logistic Regression, SVM, Random Forest Gradient boosted tree, XG Boost, and Cat Boost and we can say that light Boost is the best among the models.

II. LITERATURE SURVEY


According to the paper, From the beginning of the data mining which is used to discover new knowledges from the databases can helping various problems and helps the business for their solutions. Telecom companies improve their revenue by retaining their customers Customer churn in telecom sector is to leave a one subscription and join the other subscription in these paper they predicting the customer churn by using various R packages and they created a classification model and they train by giving him a dataset and after training they can classify the records into churn or non-churn and then they visualize the result with the help to visualization techniques. In this they are using logistic regression model and these models first train on training data after that they can test the model on test data to compute the performance measure of the classification model so we can get the various parameters like true positive rate, false positive rate and accuracy. According to, Telecom Customer churn prediction is a cost sensitive classification problem. Most of studies regard it as a general classification problem use traditional methods, that the two types of misclassification cost are equal. And, in aspect of cost sensitive classification, there are some researches focused on static cost sensitive situation. In fact, customer value of each customer is different, so misclassification cost of each sample is different. For this problem, we propose the partition cost-sensitive CART model in this paper.

Existing System

In today's technological conditions, new data are being produced by different sources in many sectors. However, it is not possible to extract the useful information hidden in these data sets, unless they are processed properly. To find out this hidden information, various analyses should be performed using data mining, which consists of numerous methods. The Churn Analysis aims to predict customers who are going to stop using a product or service among the customers. And, the customer churn analysis is a data mining-based work that will extract these possibilities. Today's competitive conditions led to numerous companies selling the same product at quite a similar service and product quality.

Proposed System

In the proposed system R programming will be used to build the model for churn prediction. It is widely used among statisticians and data miners for developing statistical software and data analysis. R is freely available and a powerful statistical analysis tool that has not yet been explored for building models for churn prediction. Churn Prediction Framework In this paper, we proposed different machine learning

| | |
|-----------------------------------------------------------------------------------|------------------------------|
|  | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

algorithms to analyse customer churn analysis. Through which we can multiple different models are employed to accurately predict those churn customers in the data set. These models are Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting Trees.

Software Requirements

- ☐ Programming Language/Platform: Python
- ☐ IDE :jupyter/pycharm

Hardware Requirements

- ☐ Processor :Intel i3 and above
- ☐ RAM :4 GB and Higher
- ☐ HardDisk :500GB Minimum

III.DIFFERENT TECHNOLOGIES FOR EXISTING ALGORITHMS

Machine learning:

Machine learning is a branch of Artificial Intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

Working of Machine Learning

- 1.A Decision Process: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabelled, your algorithm will produce an estimate about a pattern in the data.
- 2.An Error Function: An error function serves to evaluate the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
- 3.A Model Optimization Process: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model

IV. SYSTEM DESIGN

System Architecture

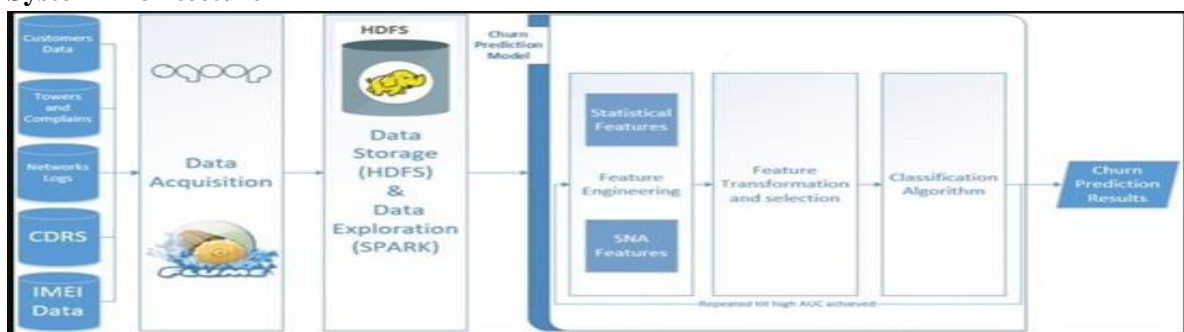



Figure 1: System Architecture

| | |
|-------------------------------------------------------------------------------------------------------------------------------------|------------------------------|
|  <p>IJESAT (Enriching the Research)</p> | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

Flow Chart

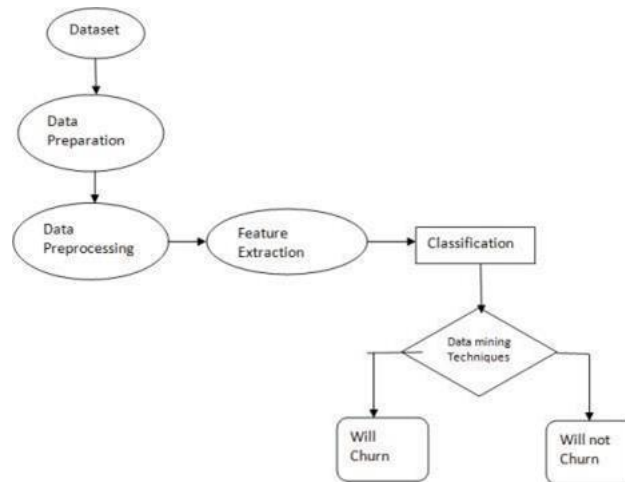


Figure 2: Flow Chart

We apply a wide range of data mining techniques from as simple as Naive Bayes, linear regression and nearest neighbours to more complex techniques as SVM, Random Forests and other ensemble methods. We solve this problem following hereunder procedure for the employee data analysis and churn prediction.


- 1) Analyse the employee dataset that consists of current and past employees' records
- 2) Clean the dataset, handle the missing information and derive new features if required
- 3) Select the features among the employee data that are suitable for the prediction of churn
- 4) Try several classifications and report the ones most suitable by comparing the accuracy, precision, recall, and F-measure results on the test data
- 5) Apply feature selection method, and select the features that are more convenient in order to predict employee churn
- 6) Build classification model
- 7) Further the prediction of churn employees on using the model
- 8) Decision on the strategies of retention

Data Selection and Cleaning

The first step in our approach is data selection and cleaning. In this step, we utilize to predict employee attrition by using HR Employee Attrition dataset provided by IBM. The dataset contains employee information such as demographics, experience, skills, nature of work or unit, position etc. This step is to identify and select the features from the employee data that are more suitable for our analysis. There were totally 35 features some of which were not useful or had same value for all records. For instance, all employees were over 18 years old. Another example, an employee ID or name may not be important and we can discard such features from the data. After removing those unnecessary features, we had 30 features. Table I shows the features of data and their type and definition.

| Dataset | Churn | Non-Churn | Total | Churn Rate |
|---------|-------|-----------|-------|------------|
| train | 160 | 640 | 800 | 0.20 |
| test | 77 | 593 | 670 | 0.11 |

Figure 3: Data set

| | |
|-----------------------------------------------------------------------------------|------------------------------|
|  | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

Data Classification Algorithms

Support vector machines (SVMs) are a non-probabilistic supervised machine learning model used for classification and regression. SVMs will train algorithms with assigned classes by separating each class through a decision boundary, also known as a hyper plane. Some problems are considered nonlinear in so far as it is difficult to draw the decision boundary. However, this can be solved by using a kernel function (also known

as a kernel trick). This function returns the dot product of the two vectors, where it then maps data points to a new, transformed, high-dimensional space. Moreover, there are several types of kernel function can be used such as linear, Gaussian, and polynomial kernel.

Random forest (RF) is one of the most powerful supervised machine learning algorithms for generating classifications and regressions. RF uses multiple decision trees to train data. Each tree votes for a classification label for a certain dataset, and then the RF model chooses which class had the most votes from the decision trees.

K-nearest neighbours (KNN) is one of the simplest machine learning algorithms and is used for both classification and regression. KNN works by specifying the value of K, which indicates the number of closest training points for a single data point. Each new data point will be classified based on the majority of votes collected from its neighbours.

Comparison of Classification Models

we compare the classification methods for understanding which method is more befitting to predict churners or non-churners. We want to figure out how accurate a classification algorithm by measuring accuracy, precision, and F-measure on the test

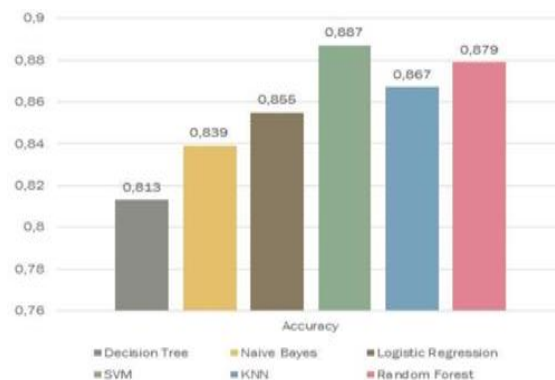



Figure 4: comparison of classification

We start with simple binary classification methods for employee churn prediction Decision Tree, Naive Bayes, and nearest neighbour's methods. Then, we try to more complex methods as Support Vector Machines (SVM), Logistic Regression, and Random Forests. We train our churn prediction models on the available labelled data. We compare our results by using different methods on the test data. The results of the basic binary predictor are in the form of churn or no churn.

Feature Selection

Feature selection techniques can be utilized to build many prediction models with different subsets of train dataset and determine those features that are and are not relevant to build reliable and accurate model. We apply feature selection method, and select the features that are more convenient in

| | |
|-------------------------------------------------------------------------------------------------------------------------------------|------------------------------|
|  <p>IJESAT (Enriching the Research)</p> | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

order to predict employee churn. In this study, we benefit from a popular feature selection method called Recursive Feature Elimination (RFE). After applying RFE method and removing redundant features, the new feature set consists of 15 features as follows: Education, Education Field, Environment Satisfaction, Gender, Job Involvement, Job Level, Job Role, Job Satisfaction, Marital Status, Over-Time, Performance Rating, Relationship Satisfaction, Stock Option Level, Years since Last Promotion, Years With Current Manager. The classification methods on the test dataset have been performed both before and after applying RFE method. The results are given in Table III and Table IV. To evaluate the performance of RFE method, the new feature set provides considerably increased accuracy and precision for almost all classification methods. On the other hand, the results show same as last time that SVM is the best method in terms of accuracy, precision and F-measure.

| Method | Accuracy | Precision | Recall | F-measure |
|---------------------|----------|-----------|--------|-----------|
| Decision Tree | 0.788 | 0.35 | 0.37 | 0.36 |
| Naive Bayes | 0.856 | 0.58 | 0.38 | 0.46 |
| Logistic Regression | 0.871 | 0.74 | 0.31 | 0.44 |
| SVM | 0.897 | 0.98 | 0.37 | 0.53 |
| KNN | 0.841 | 0.53 | 0.12 | 0.19 |
| Random Forest | 0.854 | 0.65 | 0.21 | 0.31 |

Figure 5: Data set of feature selection

Implementation

We proposed as an alternative to the user-based neighbourhood approach. We first consider the dimensions of the input and output of the neural network. In order to maximize the amount of training data we can feed to the network; we consider a training example to be a user profile (i.e. a row from the user-item matrix R) with one rating withheld. The loss of the network on that training example must be computed with respect to the single withheld rating. The consequence of this is that each individual rating in the training set corresponds to a training example, rather than each user. As we are interested in what is essentially a regression, we choose to use root mean squared error (RMSE) with respect to known ratings as our loss function. Compared to the mean absolute error, root mean squared error more heavily penalizes predictions which are further off. We reason that this is good in the context of recommender system because predicting a high rating for an item the user did not enjoy significantly impacts the quality of the recommendations. On the other hand, smaller errors in prediction likely result in recommendations that are still useful—perhaps the regression is not exactly correct, but at least the highest predicted rating are likely to be relevant to the user.

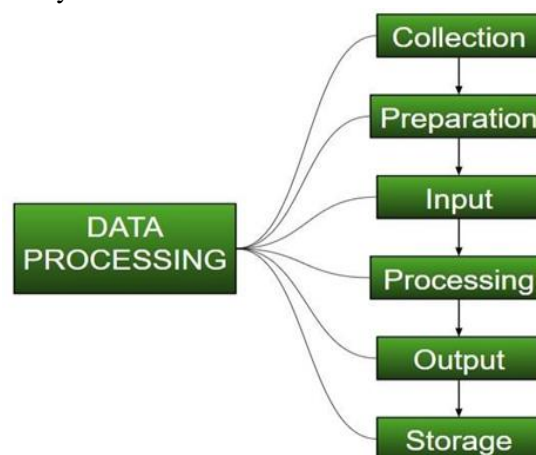



Figure 7: Data pre-processing implementation

| | |
|-----------------------------------------------------------------------------------|------------------------------|
|  | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

V. Applications of customer churn prediction

5.1 Benefits of customer churn prediction using machine learning

Identify at-risk customers

For any business that wants to enjoy the benefits of customer churn prediction, machine learning opens dozens of opportunities. Machine learning is able to analyse client behaviour and measure their probability of churning. In particular, to precisely identify churn rate, machine learning algorithms can be trained to learn the behaviour patterns of clients/partners who have already cancelled their contracts or any other relationships with a particular company and compare them with the existing ones. Then correlations between the actions of active and inactive clients are done. As a result, the algorithm recognizes the customers that are more likely to leave.

Identify pain points

Different companies lose their clients for different reasons. In most cases, there are numerous "pain points," which remain unknown for product owners. From the bad quality and absent features to unpleasant design and poor customer service — there are a lot of details which you do not take into account that your clients do. Even if your product is almost perfect, you can still reward your new customers with some attractive discounts and offers and ignore your loyal ones. When a business applies churn prediction, machine learning can do analysis and forecasts based not only on customer behaviour but also on the brand's.

Identify methods to implement

After the root cause of client churn has been identified, companies can reconsider and rebuild their products and change their business strategy accordingly. Transformed data and automated flow can be used in CRM and marketing automation systems. However, this doesn't mean that using machine learning for churn prediction is about building a certain model for a certain task. It is more about domain knowledge and an ability to deliver the best possible solution based on learning data, processes, and behaviour.


The Five Best Machine Learning Use Cases for Churn Prediction Businesses determine machine learning as a ready-made technology; however, this is not correct. Although similarities between companies operating in the same niche can be easily found, it doesn't matter they have similar business processes, flows, pricing policies.

Advantages

- Provides clarity on the business quality
- Identify whether customers are satisfied or dissatisfied with the service or product
- Compare with competitors to gauge an acceptable level of churn rate.
- Easy calculate pattern.

Disadvantages

- Doesn't provide clarity on the types of customers leaving which means you couldn't find out which one left, the new or old customer
- Doesn't differentiate companies between industry types.

| | |
|-----------------------------------------------------------------------------------|------------------------------|
|  | Open Access Research Article |
| | Volume: 23 Issue: 04 |
| | April, 2023 |

VI. Future Scope and Conclusion

Future scope

As churn prediction is one of two parts in customer churn management, and for future work, it would be interesting to investigate what features to use and how they impact churn prediction analysis. It would also be interesting to investigate, based on the variables, what measures could be taken related to retention strategies and how organizations should actively act towards customers that are predicted to churn. Another aspect, which would be interesting to further investigate, is what other methods can be used for feature selection and sampling and how they would impact the result. In addition, what 37 other algorithms that could be used and how they could be modified in order to achieve the best result.

Conclusion

In order to retain existing customers, Telecom providers need to know the reasons of churn, which can be realized through the knowledge extracted from Telecom data. In this paper, we train four machine learning models which is Logistic Regression, SVM, Random Forest, LightGBM and Gradient boosted tree and we can say that LightGBM is best in among four models and the Logistic regression and Random Forest is an average and SVM is underperforming between these models.

References

- Peng Li 1, 2, Siben Li 2, Tingting Bi 2, Yang Liu 2, " Telecom Customer Churn Prediction Method Based on Cluster Stratified Sampling Logistic Regression" in IEEE.
- Chuanqi Wang, Ruiqi Li, Peng Wang, Zonghai Chen, "Partition cost-sensitive CART based on customer value for Telecom customer churn prediction" in Proceedings of the 36th Chinese Control Conference 2017 IEEE.
- Guo-en Xia, Hui Wang, Yilin Jiang, "Application of Customer Churn Prediction Based on Weighted Selective Ensembles" in IEEE 2016.
- Rahul J. Jadhav, Usharani T. Pawar, "Churn Prediction in Telecommunication Using Data Mining Technology", in (IJACSA), Vol. 2, No.2, February 2011
- Kiran Dahiya, Surbhi Bhatia, "Customer Churn Analysis in Telecom Industry" in IEEE 2015, 978-1-4673-7231- 2/15
- N.Kamalraj, A.Malathi' " A Survey on Churn Prediction Techniques in Communication Sector" in IJCA Volume 64– No.5, February 2013
- Kiran Dahiya,KanikaTalwar, "Customer Churn Prediction in Telecommunication Industries using Data Mining Techniques- A Review" in IJARCSSE, Volume 5, Issue 4, 2015.
- R Data: <http://cran.r-project.org/>
- Data Mining in the Telecommunications Industry, Gary M. Weiss, Fordham University, USA.
- Manjit Kaur et al., 2013.Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers, IJRITCC, Volume: 1 Issue: 9
- R. Khare, D. Kaloya, C. K. Choudhary, and G. Gupta, "Employee attrition risk assessment using logistic regression analysis,".